

## ABSTRACT

Title of dissertation: PLANETARY SCALE DATA STORAGE

Benjamin Bengfort  
Doctor of Philosophy, 2018

Dissertation directed by: Professor Peter J. Keleher  
Department of Computer Science

The success of virtualization and container based application deployment has fundamentally changed computing infrastructure from dedicated hardware provisioning to on-demand, shared clouds of computational resources. One of the most interesting effects of this shift is the opportunity to localize applications in multiple geographies and support mobile users around the globe. With relatively few steps, an application and its data systems can be deployed and scaled across continents and oceans, leveraging the existing data centers of much larger cloud providers.

The novelty and ease of a global computing context means that we are closer to the advent of an Oceanstore, an Internet-like revolution in personalized, persistent data that securely travels with its users. At a planetary scale, however, a data system suffers from physical limitations that significantly impact its consistency and performance. Even with modern telecommunications technology, the latency in communication from Brazil to Japan results in noticeable synchronization delays that violate user expectations and the scale of such a system means that failure is common.

To address these issues, we explore consistency in the implementation of distributed logs, key/value databases and file systems in a planetary context. At the core of our system is hierarchical consensus, a geographically-distributed consensus algorithm that provides strong consistency, fault tolerance, durability, and adaptability to varying user access patterns. Using hierarchical consensus as a backbone, we further extend our system from data centers to edge regions using federated consistency, an adaptive consistency model that gives satellite replicas high availability at a stronger global consistency than existing weak consistency models.

In a deployment of 135 replicas in 15 geographic regions across 5 continents, we show that our implementation provides high throughput, strong consistency, and resiliency in the face of failure. Our experimental validation shows that planetary scale data systems are possible in current network environments without sacrificing consistency or performance.

# PLANETARY SCALE DATA STORAGE

by

Benjamin Bengfort

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2018

Advisory Committee:  
Professor Peter J. Keleher, Chair/Advisor  
Professor Dave Levin  
Professor Daniel Abadi  
Professor Derek C. Richardson

© Copyright by  
Benjamin Bengfort  
2018



## Preface

If needed.

## Foreword

If needed.

## Dedication

To Irena and Henry.



## Acknowledgments

I owe my gratitude to all the people who have made this thesis possible and because of whom my graduate experience has been one that I will cherish forever.

## Table of Contents

Preface	ii
Foreword	iii
Dedication	iv
Acknowledgements	v
List of Tables	vii
List of Figures	viii
List of Abbreviations	ix
1 Introduction	1
1.1 Georeplication of Data Storage . . . . .	1
2 Hierarchical Consensus	2
2.1 Overview . . . . .	2
3 Federated Consistency	3
3.1 Overview . . . . .	3
4 Adaptive Consistency	4
4.1 Overview . . . . .	4
A Previous Experiments	5
B New Experiments	6
Bibliography	7

## List of Tables

## List of Figures

## List of Abbreviations

EC	Eventual consistency
HC	Hierarchical consensus

## Chapter 1: Introduction

### 1.1 Georeplication of Data Storage

This is the first chapter. This paper will be a bit about Paxos [\[1\]](#)

## Chapter 2: Hierarchical Consensus

### 2.1 Overview

Let's start by discussing the core of my work, Hierarchical Consensus.

## Chapter 3: Federated Consistency

### 3.1 Overview

Moving on to federating consistency between strong and eventually consistent systems.



## Chapter 4: Adaptive Consistency

### 4.1 Overview

Applying machine learning techniques such as multi-armed bandits to consistency problems.

## Appendix A: Previous Experiments

Appendix A will be written here.

## Appendix B: New Experiments

Appendix B will be written here.

## Bibliography

- [1] Leslie Lamport. The part-time parliament. 16(2):133–169.